DOCUMENT RESUME

ED 070 442 LI 003 893

TITLE	Report on Information Retrieval and Library Automation Studies.
INSTITUTION	Alberta Univ., Edmonton. Dept. of Computing Science.
SPONS AGENCY	National Research Council of Canada, Cttawa (Ontario).
PUB DATE	1 Jul 72
NOTE	19p.; (34 References)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Abstracts; Automatic Indexing; Classification; Computer Science; Doctoral Theses; Information Processing; *Information Retrieval; Information Storage; *Library Automation; Masters Theses; Publications

ABSTRACT

Short abstracts of works in progress or completed in the Department of Computing Science at the University of Alberta are presented under five major headings. The five categories are: Storage and search techniques for document data bases, Automatic classification, Study of indexing and classification languages through computer manipulation of data bases, Library automation, and Information transfer processes and national networks. Faculty and student names and document titles are provided. (SJ)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 070442

REPORT ON INFORMATION RETRIEVAL AND LIBRARY AUTOMATION STUDIES

JULY 1, 1972

DEPARTMENT OF COMPUTING SCIENCE
UNIVERSITY OF ALBERTA
CANADA

LI 003 893

TABLE OF CONTENTS

1. INTRODUCTION

2. STORAGE AND SEARCH TECHNIQUES FOR DOCUMENT DATA BASES
   2.1 Coding and Storage Techniques for Retrospective Searches
   2.2 Automatic Question Modification

3. AUTOMATIC CLASSIFICATION
   3.1 Automatic Classification of Documents
   3.2 Computer Assisted Medical Diagnosis
   3.3 Holographic Image Processing

4. STUDY OF INDEXING AND CLASSIFICATION LANGUAGES THROUGH COMPUTER MANIPULATION OF DATA BASES

5. LIBRARY AUTOMATION

6. INFORMATION TRANSFER PROCESSES AND NATIONAL NETWORKS
   6.1 Small Scale Documentation
   6.2 Information Transfer
   6.3 National Policies and Education

7. STAFF AND GRADUATE STUDENTS - INFORMATION RETRIEVAL AND LIBRARY AUTOMATION

8. PUBLICATIONS

9. THESES

10. REPORTS

1. INTRODUCTION

In the Department of Computing Science at the University of Alberta a group has worked for several years in both the research and operational aspects of information retrieval and library automation. The group has cooperated with the Alberta Research Council. Computer programs have been developed and used on several document data bases for current awareness and retrospective search. Programs have been developed for on-line library automation and used on an experimental basis. Basic research has been undertaken in several areas.

The group is primarily concerned with the computing science and educational aspects of information processing. Research results have been reported in the journal literature, in M.Sc. theses, and at international meetings. The Department has graduate students engaged in information retrieval research at the M.Sc. and Ph.D. level.

Sections 2 - 6 below contain brief descriptions of work completed since 1968. In most instances each paragraph is a summary, or combination, of abstracts that appeared in the publications or reports indicated in parentheses at the end of the paragraph. The references are to the publications listed in Section 8 , to the theses of Section 9 , or to the reports included in Section 10. Any of these are available on request.

Work in progress is included at the end of each section. Further details may be obtained directly from the person concerned.

Appreciation is expressed to the National Research Council of Canada for financial support of some of the projects. Acknowledgment of support from other groups is included in the project descriptions.

2. STORAGE AND SEARCH TECHNIQUES FOR DOCUMENT DATA BASES

2.1 - Coding and Storage Techniques and Retrospective Searches

Between 1968 and 1971 computer searches were performed fortnightly on the Chemical Titles tapes as a service for current awareness. A cost function was formulated based on the number of titles searched, the length of the questions, the extent to which questions were batched, and certain details regarding the allowed forms of question. It was suggested that the search procedure should be designed to minimize computation time at the expense of convenience in the form of output, but that facilities should be included so that the user who is willing to pay the additional cost could receive output in a more convenient form. (L.H. Thiel and H.S. Heaps, Report 1968; H.S. Heaps and L.H. Thiel 1970).

An investigation was made of methods of abbreviation of English words to standard length for computer processing. Five methods of sectionalizing the data were tested on three vocabularies. The first vocabulary consisted of 63,316 title words, authors, or codens from the Chemical Titles tapes. The second vocabulary consisted of 6,354 terms from the MARC tapes. The third vocabulary consisted of 10,804 words or phrases chosen from the Dictionary of Canadian English. Abbreviation techniques were tested for the effect of inclusion of length digit, check digit, and ordering of letters selected for inclusion in the abbreviation codes. It was found that abbreviation codes may be chosen to provide a high degree of discrimination for the data bases examined. (R.L. Treleaven, Thesis 1970).

Retrospective search of large document data bases requires development of special techniques for automatic compression of data and minimization of the number of input-output operations to the computer accessible files. Also the computer program should be designed to require a relatively small amount of internal memory. A description has been given of a program that meets these requirements. The vocabulary of the data base was automatically expressed in terms of 8, 16, and 24 bit codes chosen to point to the natural spelling in a dictionary. Thus file size was reduced without the necessity for extensive processing for decoding. Use of a compressed bit string inverted index greatly reduced search time, and a storage management system enabled long strings to be processed with use of a limited amount of internal storage. Creation of "reduced" files and tables was an important feature of the program; it allowed those files

needed only by specific phases of the program to be designed to use a relatively small amount of internal storage and input-output time. (H.S. Heaps 1970a; L.H. Thiel and H.S. Heaps 1972).

Analysis has been made of the effect of using an efficient code for compression of terms within a document data base. Storage efficiency was expressed in terms of the vocabulary length and the values of certain parameters which describe the structure of the code. For vocabularies of up to 100,000 terms the average code length is approximately twelve bits. No information is lost through term truncation or abbreviation. It was shown that the tables required for coding and decoding may be ordered for rapid access without reduction in the ease of update. (H.S. Heaps 1972a).

Work in progress by E. Schuegraf is a theoretical and experimental analysis of the use of word fragments as the basic vocabulary terms for information retrieval from large document data bases. Emphasis is being placed on considerations that affect storage requirements, search times, and query language capabilities. The investigation includes selection of key fragments from a set of 1,2,3,4,5,6-character sequences, choice of a method of coding the terms, storage of the inverted index, suppression of false bits, decoding of titles, and storage requirements consistent with small search times.

Also in progress is development of an operational retrospective search service for implementation by the National Science Library. Appreciation is expressed to the National Science Library for their financial support.

2.2 - Automatic Question Modification

A program for computer retrieval of papers in acoustics was used to search titles of papers that appeared in four journals between January 1955 and December 1967. The four journals were Journal of the Acoustical Society of America, Soviet Physics-Acoustics, Acustica, and Journal of Fluid Mechanics. Title words and author names were truncated to five letters. Search questions were allowed in the form of weighted terms connected by OR logic within parameters connected by AND or NOT logic. This data base, and a similar one for computing science literature, has been used in a number of subsequent studies including an examination of associative search methods. (D.M. Heaps and H.S. Heaps 1968; H.S. Heaps 1968).

Three measures of effectiveness of an information

retrieval system were formulated in terms of a user's estimate of the relevance of items output. In each instance the type of allowed question logic was postulated without specification of certain parameters which denote the weights attached to the question terms. The parameters were then determined to maximize the search effectiveness as measured by user satisfaction. The values determined for the parameters depend on certain statistics of the data base. The search effectiveness is then optimum for the permitted form of question, the measure of output relevance, and for data bases of similar statistics. The techniques that were used are analogous to those used to define a matched filter and a Wiener root-mean-square filter. (H.S. Heaps and W.C.C. Ko 1970; W.C.C. Ko, Thesis 1970; H.S. Heaps 1971).

Work in progress by A. Lo relates to an automatic method for choice of index terms of a document data base, and assessment of the relative value of the terms. When a question is input a feedback procedure is used to successively modify the request.

3. AUTOMATIC CLASSIFICATION

3.1 - Automatic Classification of Documents

Techniques for automatic classification of documents according to subject categories have been examined for a data base of 1572 titles of papers published in the Journal of the Acoustical Society of America during 1966, 1967, 1968, and 1971. It was found that latent class analysis does not form a useful technique. Attribute analysis as proposed by Maron was found to be satisfactory with use of a proposed procedure for choice of keywords from the titles. A modified application of attribute analysis was based on maximization of correct classifications of base documents with use of not more than two keywords for the computation of joint word occurrences, but with exact values of joint occurrences instead of estimates. The classification efficiency was similar to that obtained by direct application of attribute analysis. (S. Akiyama, Thesis 1972).

Work in progress by F. Chan includes an examination of the feasibility of document classification based on application of the concept of fuzzy relations and determination of significant features.

3.2 - Computer Assisted Medical Diagnosis

The problem of automatic diagnosis by use of a computer has been expressed as an optimization problem in which parameters are chosen to minimize the diagnosis errors in reference to a previously treated set of patients. The results were expressed in terms of statistical measures of mutual associations of symptoms, and of symptoms with diseases. A decision criterion was discussed, and a formula was derived to describe the diagnostic value of each symptom. It was not necessary to make assumptions regarding mutual exclusiveness of diseases or statistical independence of symptoms. (H.S. Heaps 1972b).

Work in progress by J. Cumberbatch includes application of the above theory to analyse hospital data relating to 300 patients with respect to a disease symptom complex of six diseases and eleven symptoms. Appreciation is expressed to Dr. P.A. Scheinok who made the data available.

3.3 - Holographic Image Processing

Work in progress by D.K.K. Lam relates to pattern recognition through hologram interferometry and suitable processing of the image.

4. STUDY OF INDEXING AND CLASSIFICATION LANGUAGES THROUGH COMPUTER MANIPULATION OF DATA BASES

An experimental computer program was written for on-line interactive construction of a thesaurus. The program allowed the user to create his thesaurus and use it at the same time. Such a program could be used in a library or information centre which required a thesaurus for word control. The project was concerned with the development of an initial query language. If such an on-line thesaurus were to be run as a production system in a library, additional programs would be needed to link the thesaurus to catalog search and to circulation. The program should also be optimized. It was written in PL/1. (A.L.S. Wong and D.M. Heaps, Report 1969).

A set of experimental programs were written to manipulate MARC tapes. Three aspects were covered: 1. Programming to achieve fast code conversion from ASCII to EBCDIC and to perform word counts. 2. Programming to dump, strip, and relate fields from the MARC tape. 3. Programming to reformat the MARC tapes to search them on author and title using the programs developed at the University of Alberta for searching the Chemical Titles tapes. Tests were run and the results discussed with students in the School of Library Science. (D.M. Heaps, V. Shapiro, D. Walker, and F. Appleyard 1970).

An on-line thesaurus was developed to serve as a primary aid in classifying, indexing, and searching a specific water resource data base. Users are persons responsible for water resource management decisions. The data base contains material in bibliographic format of non-standard type. It includes research project descriptions, research grant applications, monographs, journal articles, abstracts of statutes, entire statutes, and so forth. The material is accessed and controlled, and new documents indexed and/or classified, through the on-line thesaurus. The system has a thesaurus, a data base, and a class structure (schedules). Programs and documentation are available. Programs are written in IBM360 Assembler. Appreciation is expressed to Environment Canada for financial assistance. (F. Alber and D.M. Heaps 1971; F. Alber, Thesis 1972).

The availability of machine readable data bases, and of increasingly sophisticated computer programs and methods of operation, have made possible the investigation of the indexing and classification processes through manipulation of data bases by computers. A study was carried out that used data bases such as the MARC tapes, UDC schedules, a UDC indexed data base, and a water resource thesaurus. A

methodology was developed to test the suitability of the LC and UDC classifications for control of a water resource document collection. The techniques involved the writing of programs that determined common subject headings, which were then analysed for appropriate classifications. A concordance was developed between the Water Resources Thesaurus and the UDC schedules. Appreciation is expressed to Environment Canada for financial assistance, and to FID and R. Freeman who made data bases available. (M. Mercier, G.A. Cooke, and D.M. Heaps 1971; M. Mercier, Thesis 1972).

Work in progress by S.R. Dobay relates to the efficient use of networks and requires investigation of two important problems. These are: 1. The direct matching of the resources of the network to the user. 2. The transformation or conversion of information from one level, or one vocabulary, to another. Machine-readable data bases are being used, and techniques developed, to allow concept mapping. One family of concepts is mapped on another and an intermediate lexicon produced. It is postulated that these techniques, and the lexicon, will assist classification and retrieval. The data bases used are the L.C. Subject Heading Tape and the Water Resources Thesaurus.

5. LIBRARY AUTOMATION

A preliminary report outlined a command structure for use with an on-line library automation system. The commands included ones that were user oriented, circulation oriented, and order oriented. (D.M. Heaps and H.S. Heaps, Report 1968).

A pilot study was undertaken to determine the feasibility of designing a system to automate card catalog searching in a library using the Universal Decimal Classification, in both batch and on-line mode. The use of the UDC in computer searching of the file was examined. An attempt was made to computerize a section of the card catalog as it existed and no changes were made in its basic format. No attention was given to a program for handling the circulation of books. The unique features of the UDC were only marginally explored, but the colon concatination was utilized. The study was carried out in cooperation with the Boreal Institute Library. (D.M. Heaps, L.S. Easton, E.R. Macallister, and R.L. Pallister 1970).

Some of the most serious problems in library automation arise because of the size of the files that must be stored on computer accessible devices. A discussion was presented of the techniques and approaches available for optimizing the storage requirements for bibliographic data with the use of data compression. Reference was also made to file organization and access methods. Coding and decoding problems were dealt with. (W.D. Reid and H.S. Heaps 1971).

A study of the application of on-line computer systems to library automation was divided into two parts. The first part involved the design, implementation, and evaluation of an integrated library automation system, the IT System, to encompass an on-line catalog subsystem and a real-time circulation subsystem. Implementation was in the Departmental library. The second part of the study involved the design of a computer file structure to support an on-line integrated library automation system capable of serving the needs of a large academic library. The proposed design was confined to the files necessary for the support of an on-line catalog subsystem and a real-time circulation subsystem. The design attempted to produce a file structure economical in terms of storage requirements and CPU time, and also able to provide very short response times for most on-line transactions and queries. Considerable attention was given to a new method for construction of the inverted index files of the on-line catalog subsystem. Description of the method, based on the principles of virtual hash addressing, covered the detailed structures of the index files for authors, titles, and LC call numbers, the procedures

used in generation, maintenance, and search of these files, and the theoretical performance in terms of storage requirements and file access times. (J.J. Dimsdale, Thesis 1971).

Work in progress by J.J. Dimsdale is development of an optimal design for the on-line catalog subsystem of a comprehensive on-line library automation system for a library that contains approximately one million titles. It is supposed that the total system also includes an on-line acquisition subsystem, a real-time circulation system, and possibly an on-line cataloging subsystem. It is proposed to develop a cost-time function, and to use it to provide for optimal trade-off between such factors as file storage cost, updating time and cost, and query response times.

Work in progress by J.A. Benbow is the design and implementation of an integrated on-line automated library system for use with the library of the Boreal Institute at the University of Alberta. This library contains approximately 7,000 books and more than 10,000 documents in addition to reports, periodicals, newspaper clippings, and maps. It is indexed through use of the Scott's Index variation of the standard UDC classification. The automated system will include a catalog search subsystem, a computer assisted cataloging subsystem, a real-time circulation subsystem, and an acquisition subsystem.

Also in progress is development of a set of computer programs for bibliographic processing for use in the library of the Boreal Institute. Appreciation is expressed to the Donner Canadian Foundation for their financial support.

6. INFORMATION TRANSFER PROCESSES AND NATIONAL NETWORKS

6.1 - Small Scale Documentation

A study was made of the retrieval needs of a small research group who used and produced charts. A small manual coordinate-index type system was to be controlled by the research scientists with the aid of one clerical assistant. Problems encountered were the need for instruction of the scientists in techniques of document control, the difficulties of controlling charts in every stage of production, and the requirement for control of, and personal entry into, the system by the working scientists. (D.M. Heaps, Report 1968).

Experimental work in query languages for on-line personal documentation was carried out at the University of Alberta in 1967 and 1968. The computer used was an IBM 360/67 with on-line typewriter terminals. The data base comprised the personal files of one faculty member. The investigation revealed certain fundamental differences in query languages. It was found that query languages for large scale batch systems tend to be fixed, but may be complicated. Query languages for small scale personal systems must be simple, but must allow for changes. The query languages tested allowed various forms of question logic. Basic query language instructions were grouped under three general modes, START, CHANGE, DISPLAY. It was decided that an efficient query language must direct the search, keep track of purging, allow for choice in change, telescope instructions, correct errors, and serve as its own manual of instruction. Initial programming was in APL as this was the only language then available through terminals. (D.M. Heaps and P. Sorenson 1968; D.M. Heaps and W. Harris, Report 1969).

6.2 - Information Transfer

Methods of information handling have always arisen from the historical needs of the time. It is suggested, moreover, that information has always represented power. The organization of the information and the type thought to be useful reflects the current social organization and demands. Until the late 1940's information regarded as valuable was hierarchically arranged and was predominately legal, literary, historical or political, and predominately published in monographs. This period is defined as the CLASSICAL period. With the great importance of science and technology in the 1950's classification broke down; information was needed in discrete units (reports and articles), and was then discarded. This period is defined as the MODERN period. It is postulated

that we are now entering a NEOMODERN period where both types of organizations of information are needed. Information in many forms is power. The implications of these changes were discussed. (D.M. Heaps, 1969).

A report was made on an information retrieval experiment which monitored 22 questions submitted by professional laboratory staff to an early experimental SDI service. The questions asked, answers received, analysis of relevance, and lessons learned were discussed in some detail. (G.A. Cooke and D.M. Heaps 1970).

Linking of various types of users with suitable data banks will become a problem with the growth of information networks. The documents in the data banks display characteristic styles. A system has been described whereby certain fundamental characteristics of style are recognized by the computer and reduced to a pattern output by a CALCOMP plotter. It has been proposed that such automatic "pattern recognition" will assist in directing groups of users to specific data banks. Attention has been given to the style of scientific and technical documents as these form the bases of many present day data banks. (D.M. Heaps and W. Ingram 1971).

Work in progress by W. Ingram involves further implementation and testing of the system for recognition of style.

6.3 - National Policies and Education

Almost all major industralized countries have carried out a series of studies on scientific and technical information and science policy. Canada has taken part in this process with studies largely organized by the Federal Government. The more significant Canadian studies, their influence upon and their interaction with national policy, and their implications for the education of information scientists have been outlined. The basic aim was to elicit Canadian official and non-official policy. Certain subjective judgements were expressed. (D.M. Heaps and G.A. Cooke 1970). Implications for library scientists have been discussed. (H.S. Heaps 1970b).

The state of user education in Canada has been surveyed. The inclusion of information science within a computing science university curriculum, and the operation of a laboratory to demonstrate automatic techniques, has been described. (D.M. Heaps and M.K. Pavars 1970; J. Heyworth 1971; D.M. Heaps and J. Heyworth 1971).

7. STAFF AND GRADUATE STUDENTS - INFORMATION RETRIEVAL AND LIBRARY AUTOMATION

H.S. Heaps, Professor
D.M. Heaps, Assistant Professor
L.H. Thiel, Analyst
J. Heyworth, Information Specialist

Ph.D. Students:

J. Cumberbatch
J.J. Dimsdale
D.K.K. Lam
E. Schuegraf

M.Sc. Students:

F. Chan
S.R. Dobay
W. Ingram
A. Lo

Previous Graduate Students

F.M. Alber (1969-71)
S. Akiyama (1969-72)
W.C.C. Ko (1968-70)
M.A. Mercier (1969-71)
W.D. Reid (1970-72)
R.L. Treleaven (1968-70)

8. PUBLICATIONS

(Listed Chronologically)

D.M. Heaps and H.S. Heaps, 1968. Computer retrieval of papers in acoustics. Journal of the Acoustical Society of America, vol. 43, 1461-1463.

D.M. Heaps and P. Sorenson, 1968. An on-line personal documentation system. Proceedings American Society for Information Science, vol. 5, 201-207.

H.S. Heaps, 1968. Boolean, fractional, and associative searches on truncated title words. Proceedings American Society for Information Science, vol. 5, 179-184.

D.M. Heaps, 1969. Problems and promises in information handling: the nature of information; its transfer and manipulation. Alberta Library Association Bulletin, vol. 2, 19-30.

G.A. Cooke and D.M. Heaps, 1970. An experiment in information retrieval. Chemistry in Canada, vol. 22, 17-21.

D.M. Heaps and G.A. Cooke, 1970. National policies, national networks, and national information studies in Canada. Proceedings American Society for Information Science, vol. 7, 199-203.

D.M. Heaps, L.S. Easton, E.R. Macallister, and R.L. Pallister, 1970. Automation of a UDC based library for searching purposes. Proceedings Second Seminar on UDC and Mechanized Information Systems, Frankfurt. Danish Centre for Documentation, 1971, FID 405, No. 11, 33 pp.

D.M. Heaps and M.K. Pavars, 1970. A survey of Canadian user education in science and technological sources. International Congress of Documentation, Buenos Aires, 21 pp.

D.M. Heaps, V. Shapiro, D. Walker, and F. Appleyard, 1970. Search programs for the MARC tapes at the University of Alberta. Proceedings Annual Meeting of the Western Canada Chapter of the American Society for Information Science, Vancouver, 83-93.

H.S. Heaps, 1970a. Efficient coding, storage, and search methods for information retrieval. IEEE Canadian Symposium on Communication, Montreal, 33-34.

H.S. Heaps, 1970b. The dissemination of scientific and technical information in Canada: implications for library scientists. Alberta Library Association Bulletin, vol. 2, 18-22.

H.S. Heaps and W.C.C. Ko, 1970. Automatic adaptive processing of questions in document retrieval. Proceedings American Society for Information Science, vol. 7, 319-321.

H.S. Heaps and L.H. Thiel, 1970. Optimum procedures for economic information retrieval. Information Storage and Retrieval, vol. 6, 137-153.

F. Alber and D.M. Heaps, 1971. Classifying, indexing, and searching resource management information via on-line thesaurus. Proceedings Annual Meeting of the Western Canada Chapter of the American Society for Information Science, Banff, 107-116.

D.M. Heaps and J. Heyworth, 1971. An information science option in a computing science curriculum. Proceedings Second International Conference on Training for Information Work, Rome, 156-163.

D.M. Heaps and W. Ingram, 1971. Computer recognition and graphical reproduction of patterns in scientific and technical style. Proceedings American Society for Information Science, vol. 8, 257-261.

H.S. Heaps, 1971. Criteria for optimum effectiveness of information retrieval systems. Information and Control, vol. 18, 156-167.

J. Heyworth, 1971. An information retrieval laboratory for computing sceince students. Proceedings Annual Meeting of the Western Canada Chapter of the American Society for Information Science, Banff, 59-64.

M. Mercier, G.A. Cooke, and D.M. Heaps, 1971. The study of UDC and other indexing languages through computer manipulation of machine-readable data bases. Proceedings International Symposium: UDC in Relation to Other Indexing Languages. Sponsored by Yugoslav Centre for Technical and Scientific Documentation and FID, Hercog Novi, Yugoslavia (In press).

W.D. Reid and H.S. Heaps, 1971. Compression of data for library automation. Automation in Libraries 1971, Canadian Association of College and University Libraries, 2.1-2.21.

H.S. Heaps, 1972a. Storage analysis of a compression coding for document data bases. INFOR Canadian Journal of Operational Research and Information Processing, vol. 10, 47-61.

H.S. Heaps, 1972b. A general theory for automatic diagnosis. Kybernetes, International Journal of Cybernetics and General Systems (In press).

L.H. Thiel and H.S. Heaps, 1972. Program design for retrospective searches on large data bases. Information Storage and Retrieval, vol. 8, 1-20.

9. THESES

(Listed Chronologically)

W.C.C. Ko, 1970. Some optimum criteria for information retrieval. M.Sc. Thesis, 99 pp.

R.L. Treleaven, 1970. Abbreviation of English words to standard length for computer processing. M.Sc. Thesis, 114 pp.

J.J. Dimsdale, 1971. Application of on-line computer systems to library automation. M.Sc. Thesis, 400 pp.

F.M. Alber, 1972. On-Line thesaurus design for an integrated information system. M.Sc. Thesis, 158 pp.

M. Mercier, 1972. Study of the UDC and other indexing languages through computer manipulation of machine-readable data bases. M.Sc. Thesis, 90 pp.

S. Akiyama, 1972. Automatic document classification systems. M.Sc. Thesis, 97 pp.

10. REPORTS

(Listed Chronologically)

D.M. Heaps, 1968. Chart storage and retrieval project. Atlantic Oceanographic Laboratory, Bedford Institute, Dartmouth, Nova Scotia, Report B1 68-1, 49 pp.

D.M. Heaps and H.S. Heaps, 1968. Report on a program for library automation. Department of Computing Science Publication No. 12, 25 pp.

L.H. Thiel and H.S. Heaps, 1968. Computer search of Chemical Titles at the University of Alberta. Department of Computing Science Publication No. 15, 50 pp.

D.M. Heaps and W. Harris, 1969. The development of query languages for on-line personal documentation systems. Preliminary Report. Department of Computing Science, 21 pp.

A.L.S. Wong and D.M. Heaps, 1969. 'Thesauri', an on-line program for constructing a thesaurus. Department of Computing Science Publication No. 20, 31 pp.